EXHIBIT 3

Global Epidemiology 7 (2024) 100140

Contents lists available at ScienceDirect

Global Epidemiology

journal homepage: www.sciencedirect.com/journal/global-epidemiology

Commentary

Quantitative recall bias analysis of the talc and ovarian cancer association

Julie E. Goodman *, Leon M. Espira, Ke Zu [1], Denali Boon

Gradient, One Beacon Street, 17th Floor, Boston, MA 02108, United States of America

Many epidemiology studies have evaluated the association between perineal talc use and ovarian cancer. It has been shown in several reviews and meta-analyses that, while case-control studies generally report small positive associations, the results of cohort studies examining talc use and ovarian cancer overall are consistently null [1–6]. For example, Lynch et al. [7] noted, "None of the five prospective cohort studies reported any statistically significant associations between genital talcum powder use and risk of epithelial ovarian cancer, and relative risk estimates were close to unity." They added, " Of the 26 case-control studies evaluated, 18 (69%) reported at least one statistically significant odds ratio for ovarian cancer and genital talcum powder use" [7].

Meta-analyses of these studies, driven by the case-control studies' results, have frequently reported a small but statistically significant increase in ovarian cancer risk associated with talc use (*e.g.*, [7,8]). For example, Penninkilampi and Eslick [8] reported a meta odds ratio (OR) of 1.35 (95% confidence interval [CI]: 1.27–1.43) for case-control studies, 1.06 (95% CI: 0.90–1.25) for cohort studies, and 1.31 (95% CI: 1.24–1.39) overall.

All epidemiology studies of talc and ovarian cancer of which we are aware rely on self-reported talc use. Recall error rates have been reported in several studies that used self-reported exposures (*e.g.*, [9–12]), but to the best of our knowledge, only one such study has examined recall of perineal talc use [13]. O'Brien et al. [13] evaluated data from the Sister Study, a US-based prospective cohort study of women aged 35–74 who had a sister with a history of breast cancer. When examining genital talc use in the 12 months prior to baseline, reporting consistency was 86% (95% CI: 86–87%), with 27% and 21% of participants reporting use when asked at enrollment and at follow-up, respectively. This corresponds to a recall sensitivity and specificity of 63% (95% CI: 62–64%) and 95% (95% CI: 95–95%), respectively [13]. Among women with ovarian cancer ($n = 125$), the consistency of talc use reporting was 85% (95% CI: 78–91), with 28 and 33% reporting use at enrollment and follow-up, respectively. This corresponds to a recall sensitivity and specificity of 83% (95% CI: 66–93%) and 87% (95% CI: 78–93%), respectively [14].

To better understand the potential impact of recall bias on the results of case-control studies of perineal talc use and ovarian cancer, we conducted a quantitative bias analysis using the sensitivity and specificity information reported by O'Brien et al. [13] and previously published case-control data [15] as a case study.

Methods

We used data from the largest case-control study of talc use and ovarian cancer [15] to form our study population. Cramer et al. [15] recruited 2041 ovarian cancer cases and 2100 controls from Massachusetts and New Hampshire in three phases between 1992 and 2008. The investigators reported that any genital powder use was associated with an increased risk of ovarian cancer (OR = 1.33, 95% CI: 1.16–1.52).

To assess the potential impact of recall bias, we examined seven scenarios. For Scenarios 1 and 2, we replicated the analyses reported by Cramer et al. [15], though we did not have the raw data to adjust for covariates. Scenario 1 assumes no recall bias. Scenario 2 assumes 99% sensitivity and specificity for cases' recall, and 82% sensitivity and 99% specificity for controls' recall, which is the amount of misclassification reported by the authors that would nullify the risk estimate. For Scenarios 3 and 4, we used the sensitivity and specificity reported in O'Brien et al. [13] for cases (83 and 87%, respectively, and their respective 95% CIs which were calculated using the EpiR [16] package in R v4.3.1 [17]), with 0–5% exposure misreporting among controls in Scenario 3, and 63% sensitivity and 95% specificity and their respective 95% CIs for controls in Scenario 4 (as reported for the whole cohort by O'Brien et al. [13]). For Scenarios 5–7, we assumed the sensitivities and specificities ranged between 90 and 100%, to reflect the potential for lower levels of recall bias that have been reported for other self-reported exposures [9–12].

To replicate analyses reported in Cramer et al. [15], we used single values for sensitivity and specificity to calculate ORs and CIs (Scenarios 1 and 2). For all other scenarios we used a probabilistic approach

* Corresponding author at: One Beacon Street, 17th Floor, Boston, MA 02108, United States of America.
E-mail address: jgoodman@gradientcorp.com (J.E. Goodman).
[1] Present address: Merck & Co., Inc., 33 Avenue Louis Pasteur, Boston, MA 02115, USA.

https://doi.org/10.1016/j.gloepi.2024.100140
Received 5 January 2024; Received in revised form 9 February 2024; Accepted 6 March 2024
Available online 11 March 2024

incorporating random and differential errors in self-reported exposures using the *episensr* [18] package in R v4.3.1 [17].

For analyses using data from O'Brien et al. [13] (Scenarios 3 and 4), we assumed triangular distributions with the point estimates for sensitivity and specificity as the mode and the lower and upper limit of the sensitivity or specificity 95% CI as the minimum and maximum values of the distribution. We chose a triangular distribution because it is "an ideal distribution when the only data on hand are the maximum and minimum values, and the most likely outcome" [19]. The reported specificity 95% CI among all participants in O'Brien et al. [13] was 95–95%. In order to run a probabilistic model, we used 94.9 and 95.1%, as the minimum and maximum values, respectively, in Scenario 4.

For Scenarios 5–7, we modeled results for a lower level of recall bias, and assumed uniform distributions because there is no indication that any one value is more likely than any other in this range. We ran 200 simulations to sample a range of sensitivities and specificities within the specified ranges and calculated ORs and simulation intervals (SIs).

Results

Although we were not able to adjust for covariates or assess the impact of potential confounding, we calculated a similar risk estimate for perineal talc use and ovarian cancer as Cramer et al. [15] (Table 1, Scenario 1), and confirmed that the results would be null if controls' recall sensitivity and specificity were 82 and 99%, respectively, and cases' recall sensitivity and specificity were both 99% (Scenario 2). Applying the sensitivities and specificities reported in O'Brien et al. [13] resulted in a null risk estimate in Scenario 3 and reversed the direction of the association in Scenario 4 (OR = 0.62, 95% CI: 0.36–0.95). We found that if cases' and controls' recall sensitivity and controls' recall specificity ranged from 95 to 100%, and if cases' recall specificity ranged from 90 to 95% or 90 to 100%, the ORs were attenuated, with 95% SIs that span the null (Scenarios 5 and 6). However, when we assumed non-differential recall sensitivity and specificity ranges of 95–100% for both cases and controls, the resulting risk estimate was very similar to the original estimate reported by Cramer et al. [15] (Scenario 7). None of these scenarios account for confounding or other potential biases, such as selection bias, which also may have impacted estimated risks.

Discussion

In all of the published case-control studies on talc and ovarian cancer of which we are aware, exposure is determined based solely on participants' self-reported prior talc use, sometimes many years or decades in the past. Of these studies, only Cramer et al. [15] conducted a recall bias analysis, and did so using point estimates of sensitivity and specificity, rather than distributions. The investigators recalculated the OR assuming 99% recall specificity and sensitivity in cases and 99 and 82% recall specificity and sensitivity in controls. Using these assumptions, the OR was 1.00. However, the assumptions of 99% recall specificity and sensitivity are potentially unrealistically high, based on real world data reported in O'Brien et al. [13], and the Cramer et al. [15] analysis did not account for uncertainty in these very low assumed misclassification rates or total uncertainty in the measured OR of 1.33. Our results demonstrate that recall bias could have potentially influenced the observed association.

In all but two of our recall bias scenarios, the observed association between talc use and ovarian cancer was attenuated, with CIs/SIs spanning the null. In one scenario, based on data from the Sister Study, the OR was <1 and statistically significant (Scenario 4, Table 1). We do not think this reflects a protective role of talc, as there are still likely other biases and uncontrolled confounding present, but rather demonstrates that recall bias can drastically impact results. The one instance in which the OR remained statistically significantly >1 was when ranges of sensitivity and specificity of 95–100% for both cases and controls were modeled (Scenario 7, Table 1). We believe this scenario is less likely than some of the others, as it reflects non-differential misclassification, but included it for completeness.

There are examples of other studied associations in environmental epidemiology research, such as brain cancer and cell phone use, for which numerous case-control studies reported increased risks, but most prospective cohort studies did not confirm these positive findings [20]. The most recent reviews on this topic have concluded that the null findings of cohort studies are likely to be accurate [20] and that cases in case-control studies were more likely to over-report past cell phone use, thereby biasing results towards false-positive findings [21].

With respect to talc, Schildkraut et al. [22] hypothesized that the publicity of two high-profile lawsuits in 2014 may have influenced study participants' recollections of talc use. Schildkraut et al. [22] found the OR for genital talc use was 1.19 (95% CI: 0.87–1.63) for women interviewed prior to 2014 and 2.91 (95% CI: 1.70–4.97) for women interviewed in 2014 or later. This provides additional support for recall bias among more recent case-control studies.

Our research question was purposefully narrow in scope, focused on quantifying the potential impact of recall bias of talc use in case-control studies of ovarian cancer. This was motivated by reviews and meta-

Table 1
Quantitative Recall Bias Analysis of Perineal Talc Use and Ovarian Cancer Association in the Cramer et al. [15] Study

	Cases[a] (n = 2041)		Controls[a] (n = 2100)		Bias-Adjusted OR[b]	95% CI or *SI*[c]
	Sensitivity (%)	Specificity (%)	Sensitivity (%)	Specificity (%)		
Cramer et al. [15]	100	100	100	100	**1.33**	**1.16–1.52**
Scenario 1	100	100	100	100	**1.30**	**1.13–1.48**
Scenario 2	99	99	82	99	1.00	0.87–1.16
Scenario 3[d]	83 (66–93)	87 (78–93)	97.5 (95–100)	97.5 (95–100)	1.07	0.67–1.64
Scenario 4[d]	83 (66–93)	87 (78–93)	63 (62–64)	95 (94.9–95.1)	**0.62**	**0.36–0.95**
Scenario 5[e]	95–100	90–95	95–100	95–100	1.10	0.94–1.28
Scenario 6[e]	95–100	90–100	95–100	95–100	1.21	0.96–1.47
Scenario 7[e]	95–100	95–100	95–100	95–100	**1.33**	**1.13–1.54**

Notes
CI = Confidence Interval; OR = Odds Ratio; SI = Simulation Interval.
Bold = Statistically Significant.
(a) Cases and controls were recruited from Massachusetts and New Hampshire in three phases between 1992 and 2008.
(b) Adjusted for several covariates by Cramer et al. [15], but not in Scenarios 1–7 because covariate data were not available.
(c) 95% CI as reported by Cramer et al. [15], and in Scenarios 1-2 in which single values were used as sensitivity and specificity inputs, and 95% SI for Scenarios 3–7, in which a range of input values were used in the simulations.
(d) Triangular probability distributions: mode (minimum, maximum).
(e) Uniform probability distributions within each range reported.

analyses reporting different results by study design (*e.g.*, [7,8]) and the recent availability of data on talc recall in the prospective Sister's Study [13]. Realistically, the impact of all sources of bias and confounding are likely complex. However, while we could not account for confounders in our analysis because we did not have access to the raw data from Cramer et al. [15], the multivariate-adjusted OR reported by Cramer et al. [15] is similar to the crude OR we calculated. This indicates that confounding is not likely a large issue when comparing our results to those of the original study, although uncontrolled and residual confounding cannot be completely ruled out for either our analysis or the original study.

Other limitations of our analysis include that the number of ovarian cancer cases in the Sister Study population was small and loss to follow-up was high [14]. In addition, because study participants have a sister with a history of breast cancer, results from this study may not be generalizable. Because of these issues, we also conducted analyses assuming plausible ranges of recall sensitivities and specificities in cases and controls, based on studies of other self-reported exposures that were validated, and found that in most cases, risks were still attenuated. Finally, while we focus on recall bias here, other sources of bias may have also impacted results in this and other epidemiology studies of talc use and ovarian cancer, both towards and away from the null [23–27].

Although cohort studies can be subject to other types of bias, prospective exposure assessments cannot be influenced by case status. Cohort studies have consistently reported no overall association and no exposure-response relationship between perineal talc use and ovarian cancer risk overall. Also, experimental studies do not support a causal association [7,28]. Our analysis, using data from the Cramer et al. [15] case-control study, demonstrates that recall bias alone may have a large impact on risk estimates. In this case, most scenarios demonstrate that recall bias results in a bias away from the null. It is likely that this bias has affected other case-control studies in a similar manner, because they have all used similar methods to estimate exposure. Additional studies examining the consistency of self-reported perineal talc use in other populations will help inform researchers and regulators on the true impact of recall bias. Recall bias should always be assessed when there is a discrepancy between case-control study results and results in cohort and animal studies when integrating these streams of evidence for causal evaluations.

Data

Data were obtained from the Cramer et al. [15] and O'Brien et al. [13] publications and personal communications with Katie O'Brien from the National Institute of Environmental Health Sciences [14]. Computing code is available in Appendix A.

CRediT authorship contribution statement

Julie E. Goodman: Writing – review & editing, Writing – original draft, Methodology, Conceptualization. **Leon M. Espira:** Formal analysis. **Ke Zu:** Writing – review & editing, Methodology, Formal analysis. **Denali Boon:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: All authors were employed by Gradient, a private environmental consulting firm, during the drafting of this manuscript. Gradient has conducted work on talc in general, and this topic specifically, in the context of litigation and regulatory comments to Health Canada. That work informed this analysis but did not influence the work presented here. The Cosmetics Alliance Canada (CA Canada) and Essential Minerals Association (EMA) (formerly Industrial Minerals Association – North America) provided funding for some initial research, but were not involved with the conception or drafting of this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gloepi.2024.100140.

References

[1] O'Brien KM, Tworoger SS, Harris HR, Anderson GL, Weinberg CR, Trabert B, et al. Association of powder use in the genital area with risk of ovarian cancer. JAMA 2020;323(1):49–59. https://doi.org/10.1001/jama.2019.20079.
[2] Gertig DM, Hunter DJ, Cramer DW, Colditz GA, Speizer FE, Willett WC, et al. Prospective study of talc use and ovarian cancer. J Natl Cancer Inst 2000;92(3):249–52. https://doi.org/10.1093/jnci/92.3.249.
[3] Houghton SC, Reeves KW, Hankinson SE, Crawford L, Lane D, Wactawski-Wende J, et al. Perineal powder use and risk of ovarian cancer. J Natl Cancer Inst 2014;106(9). https://doi.org/10.1093/jnci/dju208. dju208.
[4] Gonzalez NL, O'Brien KM, D'Aloisio AA, Sandler DP, Weinberg CR. Douching talc use and risk of ovarian cancer. Epidemiology 2016. https://doi.org/10.1097/EDE.0000000000000528.
[5] Urban N, Hawley S, Janes H, Karlan BY, Berg CD, Drescher CW, et al. Identifying post-menopausal women at elevated risk for epithelial ovarian cancer. Gynecol Oncol 2015;139(2):253–60. https://doi.org/10.1016/j.ygyno.2015.08.024.
[6] Gates MA, Rosner BA, Hecht JL, Tworoger SS. Risk factors for epithelial ovarian cancer by histologic subtype. Am J Epidemiol 2010;171(1):45–53. https://doi.org/10.1093/aje/kwp314.
[7] Lynch HN, Lauer DJ, Leleck OM, Freid RD, Collins J, Chen K, et al. Systematic review of the association between talc and female reproductive tract cancers. Front Toxicol 2023;5:1157761. https://doi.org/10.3389/ftox.2023.1157761.
[8] Penninkilampi KJR, Eslick GD. Perineal talc use and ovarian cancer: A systematic review and meta-analysis. Epidemiology 2018;29:41–9. https://doi.org/10.1097/EDE.0000000000000745.
[9] Jasti S, Siega-Riz AM, Cogswell ME, Hartzema AG. Correction for errors in measuring adherence to prenatal multivitamin/mineral supplement use among low-income women. J Nutr 2006;136:479–83. https://doi.org/10.1093/jn/136.2.479.
[10] Burton A, Wilson S, Gillies AJ. Folic acid: is self reported use of supplements accurate? J Epidemiol Community Health 2001;55:841–2. https://doi.org/10.1136/jech.55.11.841.
[11] Johnson CY, Flanders WD, Strickland MJ, Honein MA, Howards PP. Potential sensitivity of bias analysis results to incorrect assumptions of nondifferential or differential binary exposure misclassification. Epidemiology 2014;25:902–9. https://doi.org/10.1097/EDE.0000000000000166.
[12] Keil AP, Daniels JL, Hertz-Picciotto I. Autism spectrum disorder, flea and tick medication, and adjustments for exposure misclassification: the CHARGE (CHildhood Autism Risks from Genetics and Environment) case-control study. Environ Health 2014;13:3. https://doi.org/10.1186/1476-069X-13-3.
[13] O'Brien KM, Ogunsina K, Wentzensen N, Sandler DP. Douching and genital talc use: patterns of use and reliability of self-reported exposure. Epidemiology 2023;34(3):376–84. https://doi.org/10.1097/EDE.0000000000001589.
[14] O'Brien K. [National Institute of Environmental Health Sciences]. Personal communication. 2023.
[15] Cramer DW, Vitonis AF, Terry KL, Welch WR, Titus LJ. The association between talc use and ovarian cancer: A retrospective case-control study in two US states. Epidemiology 2016;27:334–46. https://doi.org/10.1097/EDE.0000000000000434.
[16] Stevenson M, Sergeant E, Firestone S. epiR: Tools for the Analysis of Epidemiological Data (R Package Version 2.0.67). https://cran.r-project.org/web/packages/epiR/index.html.
[17] The R Foundation. R Project for Statistical Computing. R (Version 4.3.1). https://www.r-project.org; 2023. Published June 6, 2023. Accessed August 25.
[18] Haine D. The episensr package: Basic sensitivity analysis of epidemiological results (R package version 1.2.0). https://dhaine.github.io/episensr; 2023. Published 2023. Accessed August 25.
[19] Elsevier BV. Triangular distribution. https://www.sciencedirect.com/topics/mathematics/triangular-distribution; 2024. Accessed: February 9, 2024.
[20] Swerdlow AJ, Feychting M, Green AC, Kheifets L, Savitz DA. Mobile phones, brain tumors, and the interphone study: where are we now? Environ Health Perspect 2011;119:1534–8.
[21] Vrijheid M, Cardis E, Armstrong BK, Auvinen A, Berg G, Blaasaas KG, et al. Validation of short term recall of mobile phone use for the interphone study. Occup Environ Med 2006;63:237–43.
[22] Schildkraut JM, Abbott SE, Alberg AJ, Bandera EV, Barnholtz-Sloan JS, Bondy ML, et al. Association between body powder use and ovarian Cancer: the African American Cancer Epidemiology Study (AACES). Cancer Epidemiol Biomarkers Prev 2016;25(10):1411–7. https://doi.org/10.1158/1055-9965.EPI-15-1281.

[23] Jurek AM, Greenland S, Maldonado G. How far from non-differential does exposure or disease misclassification have to be to bias measures of association away from the null? Int J Epidemiol 2008;37(2):382–5.
[24] Sorahan T, Gilthorpe MS. Non-differential misclassification of exposure always leads to an underestimate of risk: an incorrect conclusion. Occup Environ Med 1994;51(12):839–40.
[25] Dosemeci M, Wacholder S, Lubin JH. Does nondifferential misclassification of exposure always bias a true effect toward the null value? Am J Epidemiol 1990;132 (4):746–8.
[26] Van Smeden M, Lash TL, Groenwold RHH. Reflection on modern methods: five myths about measurement error in epidemiological research. Int J Epidemiol 2020; 49(1):338–47. https://doi.org/10.1093/ije/dyz251.
[27] Flegal KM, Keyl PM, Nieto FJ. Differential misclassification arising from nondifferential errors in exposure measurement. Am J Epidemiol 1991;134(10): 1233–44.
[28] Goodman JE, Kerper LE, Prueitt RL, Marsh CM. A critical review of talc and ovarian cancer. J Toxicol Environ Health B 2020;23:183–213. https://doi.org/10.1080/ 10937404.2020.1755402.